PEANUT BUD ORIENTATION DETECTION METHOD BASED ON FSL-YOLO
/
基于 FSL-YOLO 的花生芽向检测方法

Zhenghao LI[1], Mei WANG[2], Chunwang DONG[4], Huawei YANG[2], Zhiwei CHEN[4], Yulong CHEN *[1,3]

[1] School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo/China
[2] Shandong Academy of Agricultural Machinery Sciences, Jinan/China
[3] Institute of Modern Agricultural Equipment，Shandong University of Technology, Zibo/China
[4] Tea Research Institute of Shandong Academy of Agricultural Sciences, Jinan/China
Corresponding author: Yulong Chen;
Tel: +86-15315206471; E-mail: cyl06471@sdut.edu.cn

DOI: https://doi.org/10.35633/inmateh-78-20

Keywords: *Peanut Sprout, Small-target detection, YOLOv8, Lightweight detection head*

ABSTRACT

Peanut sprout orientation detection is a critical step in achieving automated production. However, the small target size and dense distribution of peanut sprouts in plug trays impose higher requirements on the model's ability to extract features from small objects and discriminate in densely populated scenes. In addition, the limited computational resources of embedded devices restrict the deployment of complex models. To address these challenges, this study proposes a lightweight peanut sprout orientation detection model named FSL-YOLO, based on YOLOv8. The proposed model introduces improvements in four main aspects. First, a Fast-CA module, integrating Coordinate Attention with FasterNet, is incorporated into the backbone network to enhance the perception of dense small targets while reducing the number of parameters and computational cost. Second, a lightweight downsampling module (LWDS) is designed to replace traditional convolution operations, further improving detection performance. Third, spatial and channel reconstruction convolution (SCConv) is introduced into the neck network to optimize the C2f module, thereby enhancing feature representation capability and model robustness. Fourth, an efficient lightweight detection head, Detect-L, is constructed to further reduce the model size. Experimental results demonstrate that FSL-YOLO achieves both high accuracy and lightweight performance. The model attains an mAP50 of 96.1%, representing a 2.4% improvement over the original YOLOv8, while reducing floating-point operations (FLOPs) by 51.9% and the number of parameters (Params) by 50%. These results indicate that the proposed model effectively balances detection accuracy and computational efficiency, providing a solid technical foundation for the implementation of automated peanut sprout production systems.

摘要

花生芽朝向检测是实现其自动化生产的关键环节。然而，穴盘环境中的花生芽具有目标尺寸小、分布密集的特点，对模型的小目标特征提取与密集场景判别能力提出了更高要求，同时受限的设备算力也制约了复杂模型的部署。为此，本研究提出了一种基于 YOLOv8 的轻量级花生芽朝向检测模型 FSL-YOLO。该模型主要从四个方面进行改进：在主干网络中引入融合 Coordinate Attention 与 FasterNet 的 Fast-CA 模块，以增强对密集小目标的感知并降低参数量；设计轻量级下采样模块（LWDS）替代传统卷积，进一步提高检测精度；在颈部网络采用空间与通道重构卷积（SCConv）优化 C2f 模块，提升特征表达与模型鲁棒性；构建高效轻量检测头 Detect-L，进一步压缩模型规模。实验结果表明，FSL-YOLO 在精度与轻量化方面均表现优异：mAP50 达到 96.1%，较原 YOLOv8 提升 2.4%；浮点运算量（FLOPs）降低 51.9%，参数量（Params）减少 50%。该模型兼具高精度与低复杂度的优势，为花生芽自动化生产系统的实现提供了技术基础。

INTRODUCTION

As a popular green and healthy food, peanut sprouts are rich in protein, vitamins, minerals and bioactive substances *(Wang et al., 2022)* and have both edible and health value. In recent years, its market demand has shown a rapid growth trend *(Wang et al., 2021)*. In the production process of peanut sprouts, regulating the bud orientation is the key link that affects the quality of finished products and production efficiency. If the direction of the bud is different, it will not only delay its growth process, but also easily lead to morphological bending and even deformity, which will directly affect the appearance uniformity and commodity quality of the product, and then restrict the improvement of the overall production capacity.

At present, the classification, recognition and sorting of peanut bud orientation in the plug still mainly rely on manual work. Due to the small size and dense distribution of peanut buds, the current method of relying on manual orientation recognition faces many challenges such as high labor intensity, low efficiency, high cost and labor shortage, which has become the core bottleneck restricting the industrial upgrading to scale and automation *(Chen et al., 2017)*. Therefore, the realization of rapid, accurate and automatic detection of peanut bud orientation is not only the key measure to break through the current production bottleneck, but also has important practical significance for improving the overall production capacity, reducing labor costs and promoting the intelligent transformation of the industry.

Previously, the bud classification and detection methods in the field of agriculture can be divided into the methods based on traditional image processing and the methods based on deep learning. The core of the traditional bud direction detection method is multi technology fusion. The shape and spatial information are obtained by sensors, and then the direction parameters are analyzed by image processing technology. The combination of the two can realize the accurate detection of bud direction. For example, *Wang et al*., *(2021),* proposed a non-destructive monitoring method for the growth of cucumber plug seedlings based on Kinect camera. By matching the color image and depth image of cucumber plug seedlings and other image processing methods, the germination rate and plant height parameters of cucumber plug seedlings were monitored. *Tu et al*., *(2022)*, and others designed an automatic monitoring system for rape seed germination rate under dark light environment, which greatly improved the working efficiency of measuring seed germination rate under dark light environment. However, the above-mentioned traditional image processing methods usually rely on the manually designed feature extraction rules, and the recognition performance is easy to be significantly reduced in the complex scenes with dense distribution, overlapping or significant morphological differences of peanut buds. In addition, this kind of method is sensitive to environmental changes and has limited generalization ability, so it is difficult to be stable and suitable for the changing lighting conditions in actual production. At the same time, its multi-step image processing process usually has low computational efficiency, which is difficult to meet the real-time detection requirements in large-scale and pipelined production scenes.

With the development of computer vision and deep learning technology, image recognition method based on deep learning has gradually become a research hotspot (*Gu et al*., *2024)*. Various algorithms based on convolutional neural networks have been gradually applied in agriculture. YOLO series is an important branch of object detection network and has shown great potential. For example, *Wang et al., (2022),* proposed a weed recognition method in the corn field based on the shift window Transformer network, which realized the real-time accurate recognition and fine segmentation of corn and weeds, with an average pixel accuracy of 97.18%. *Genze et al., (2022),* explored the technology path based on machine learning, and realized the germination detection, prediction and quality evaluation of corn, rye and millet, which verified the feasibility of deep learning in the monitoring of crop growth status. *Hou et al., (2021),* proposed the improved YOLOv3 model for rapid identification and orientation determination of ginger seed buds, combined with data enhancement and Diou frame regression loss function to improve the regression effect of seed bud identification, with an average accuracy of 98.2%. *Bai et al., (2023),* built an automatic detection model of wheat seed germination based on the YOLOv5 model, integrating machine vision technology and deep learning algorithm, which effectively improved the detection efficiency. *Wang et al., (2024),* based on the YOLOv5 model, introduced the attention mechanism and lightweight convolution method group random convolution to achieve small target tea bud detection under complex background. *Zhang et al., (2023)* and others proposed the improved YOLOv7 model for potato seed potato bud eye detection, which enhanced the response of bud eye characteristics and suppressed background interference by introducing attention mechanism and multi-scale fusion. *Zhang et al., (2023),* based on the RGB-D image of grape multiple bud recognition and positioning method, realized the detection of sprouts in the distant view and the recognition of secondary buds in the close-up juxtaposed bud mapping area, which improved the recognition accuracy of small target objects in the field. *Ding et al., (2025),* introduced Focaler_mpdiou into the YOLOv8 model and distilled the knowledge to improve the effect of small target detection and reduce the complexity of the model; *Li et al., (2024),* introduced partial convolution and weighted bidirectional feature pyramid network on the basis of YOLOv8 model to improve the model detection efficiency. *Ma et al., (2024),* based on the YOLOv8 model, used the PP-LCNet model and the deep separable convolution to optimize the backbone network, and introduced the global attention mechanism and the content aware reassembly of features module to further reduce the amount of model parameters and ultimately achieve model lightweight.

Deep learning technology has important application value in the field of bud detection. The detection model represented by YOLO (*Redmon et al., 2016)* series provides an effective technical basis for peanut bud

recognition through loss function optimization, attention mechanism enhancement and lightweight structure design. According to the actual needs of peanut bud orientation classification and detection in the plug environment, the task faces two challenges. First, the bud size is small and densely distributed, which requires the model to have more refined small target feature extraction ability and higher dense scene discrimination performance. Second, the computing power of equipment in the actual agricultural deployment environment is usually limited, and the complex model structure is difficult to meet the requirements of real-time and low power consumption. Therefore, this study takes YOLOv8 as the baseline network, and proposes a lightweight peanut bud detection model for small target dense scenes, aiming to take into account the detection accuracy and reasoning efficiency, so as to promote the effective application of this technology in the actual production environment.

MATERIALS AND METHODS

Image acquisition and dataset construction

In October 2024, this study collected peanut bud image data in Jinan, Shandong Province, and constructed a special peanut bud detection data set. Peanut buds of "Huayu 25" cultivated for 3 days were selected as test samples, and images were taken at a height of 0.5 meters perpendicular to the plug. All images were saved in JPEG format with a resolution of 3072 × 4096. After data cleaning and eliminating blurring, overexposure and other low-quality images, a total of 600 effective images were obtained. The image was further clipped to remove the redundant background around the plug. Finally, the data set was divided into training set, verification set and test set according to the ratio of 8:1:1.

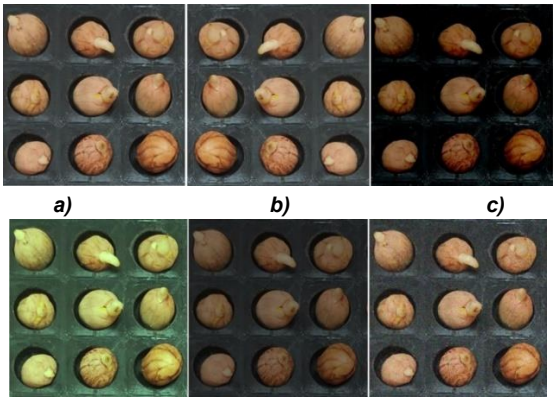
Under the constraint of the structure of the cavity in the plug, the growth direction of peanut bud mainly shows two discrete states: upward or downward. In this study, LabelMe software was used to mark the collected images manually, and the samples were divided into two types of labels according to the direction of the bud: " TRUE " indicates that the bud is facing up, and " FALSE " indicates that the bud is facing down. The annotated image is shown in Fig. 1.

a) *b)* *c)*

Fig. 1 – Peanut sprout status and dataset annotation example

a) Bud up; b) Bud downward; c) Dataset annotation

In order to improve the performance and generalization ability of the model and prevent over fitting of the model, data enhancement is used to expand the data set. According to the actual situation such as the fluctuation of lighting conditions, the change of equipment shooting angle and noise interference in the production scene, the data set is expanded by horizontal flipping, reducing brightness, changing contrast and saturation, and adding Gaussian noise. The effect of data expansion is shown in Fig. 2. After augmentation, the total number of images in the data set was expanded to 3600, including 2880 images in the training set, 360 images in the validation set and 360 images in the test set.



a) *b)* *c)*

d) *e)* *f)*

Fig. 2 – Data augmentation

a) Original image; b) Horizontal flip; c) Random brightness; d) Change saturation; e) Change contrast; f) Add Gaussian noise

Detection model based on improved FSL-YOLO

In recent years, YOLO series models have been widely used in the field of target detection because of their advantages of high precision and fast reasoning. Compared with R-CNN (*Girshick et al., 2014*), Fast R-CNN (*Ren et al., 2017*) and other two-stage algorithms, YOLO series has significant real-time advantages. Among them, YOLO8 further optimizes the balance between speed and accuracy by introducing C2f module, PANet structure and no anchor frame detection head, making it more suitable for real-time detection tasks in multi-scale complex scenes.

In order to meet the challenges of small targets, dense distribution and limited computing power of equipment in peanut bud orientation detection, this study proposes a lightweight detection model based on YOLOv8, which aims to take into account the detection accuracy and computational efficiency, and adapt to the actual agricultural application scenarios. Aiming at the task of peanut bud detection, this study proposed a set of lightweight improvement scheme of the system: introducing FasterNet Block (*Zhang et al., 2024*) and coordinate attention (*Bao et al., 2023*) in the backbone network to improve the efficiency of small target feature extraction; A lightweight downsampling module LWDS is designed to replace the traditional convolution to balance the multi-scale information retention and computational cost. The SCConv module is embedded in the neck to enhance the robustness against occlusion and light interference *(Li et al., 2023*), and an efficient detection head, Detect-L, is constructed. By concatenating depthwise convolution (*Fang et al., 2025)*, partial convolution (*Li et al., 2024)* and SimAM (*Yin et al., 2024*), the detection accuracy is maintained while reducing the computational overhead. The overall design not only ensures the lightweight of the model, but also significantly improves the recognition performance in the dense small target scene.

The network structure of the improved model is shown in Fig. 3.

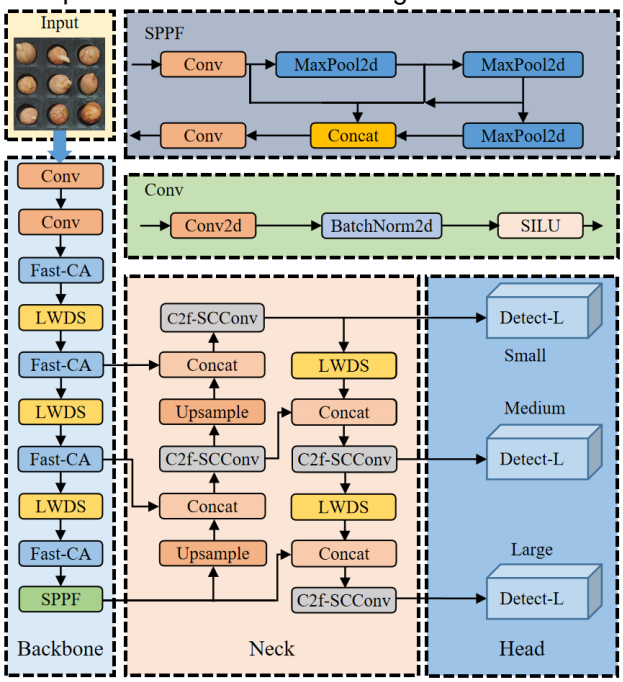


Fig. 3 – FSL-YOLO Network Structure Diagram

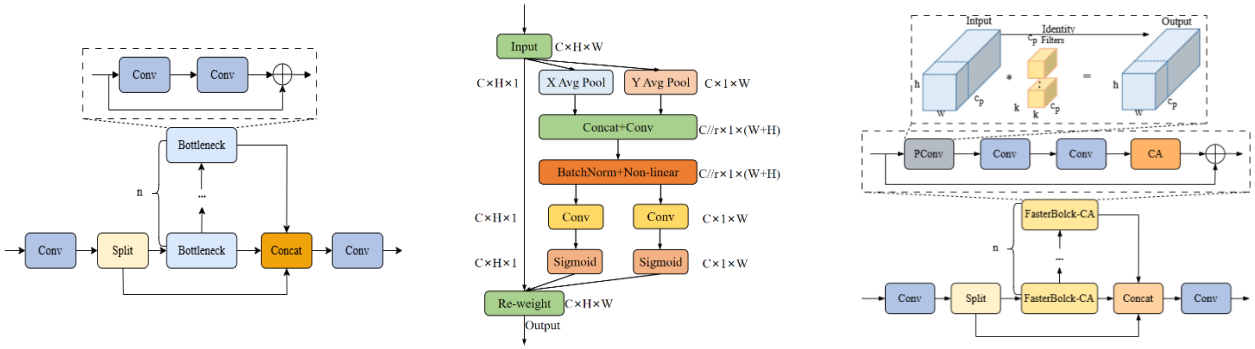
Fast-CA module

In YOLOv8, the C2f module, as the core feature extraction unit, adopts the design of "channel doubling - feature shunting - Cross stage fusion": first, the number of channels is expanded by 1 × 1 convolution to enhance the feature expression, and then the features are divided into two branches, one is directly retained, the other is deeply extracted by the Bottleneck module, and finally the two features are spliced and the channels are adjusted by convolution to achieve cross stage fusion. The structure is shown in Fig. 4. However, the design still has some limitations: on the one hand, its stacked bottleneck structure is easy to introduce redundant computing and reduce resource utilization; on the other hand, the module lacks an effective attention mechanism and is difficult to focus on key areas adaptively in scenes with dense targets, occlusion or complex background, which affects the detection accuracy.

In order to reduce the computational burden of the model and improve the reasoning speed, this study introduces the FasterNet Block, the core component of FasterNet (*Wang et al., 2024)*. The module uses partial convolution (PConv) to efficiently extract spatial features, combines point by point convolution (PConv) to fuse channel information, and forms a lightweight feature processing unit through normalization and activation

functions, which can significantly reduce the number of parameters and computational complexity, while maintaining strong feature expression ability.

In order to further enhance the ability of the model to focus on key features, a lightweight coordinate attention mechanism (CA) is introduced. The mechanism generates spatial attention weights by globally encoding the coordinate directions of the feature map, so as to adaptively strengthen important regions and suppress redundant information. Its structure is shown in Fig. 4. CA improves the sensitivity and overall robustness of the model to small target features with low computational cost.



a) *b)* *c)*

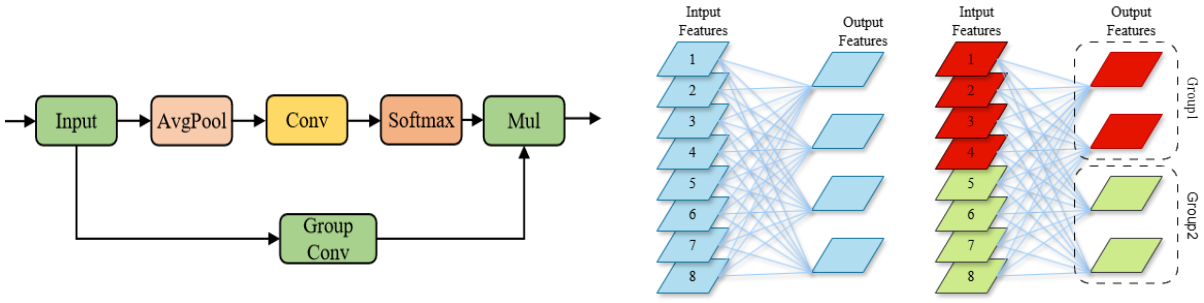
Fig. 4 – C2f and Fast CA module structure diagram

a) C2f module; b) Coordinate attention; C) Fast-CA module;

In this study, a lightweight Fast-CA module was constructed by replacing the Bottleneck structure in the C2f module with the FasterNet Block and embedding the coordinate attention mechanism (CA). While significantly reducing the computational complexity and the number of parameters, the module enhances the ability to focus on the key features of small targets through coordinate attention, and improves the efficiency of cross stage feature fusion, so as to effectively improve the detection accuracy and deployment applicability of the model in dense scenes.

LWDS module

In YOLOv8, Standard convolution is used for downsampling in the baseline design, which can extract local features, but its receptive field is limited and the calculation method is fixed, so it is difficult to effectively retain the details of small targets in complex background. Therefore, a lightweight down sampling module (LWDS) is designed to replace the original conv module. Its structure is shown in Fig. 5.



a) *b)* *c)*

Fig. 5 – LWDS Module Architecture and Convolution Operation Comparison

a) LWDS module; b) Standard convolution; c) Grouped convolution

The module uses a dual branch parallel architecture to take into account the global context and local feature extraction. The upper branch captures the overall distribution of the feature map through the global average pooling, and then generates the spatial attention weight through 1 × 1 convolution and Softmax function, so as to highlight the key areas and suppress the background interference. The lower branch uses block convolution for feature down sampling, which can significantly reduce the number of parameters while maintaining the ability of multi-scale feature extraction. Finally, the weighted fusion output of the two branch features can reduce the computational complexity and enhance the identification ability of small target features. As shown in Fig. 5, packet convolution can significantly reduce the parameters and computational complexity by grouping the input channels and independently operating with the corresponding convolution kernel, while ordinary convolution requires dense connection in the whole channel.

Through the above design, the LWDS module not only ensures the subsampling efficiency, but also significantly improves the robustness of feature extraction in dense small target scenes.

C2f-SCConv module

In order to further optimize the multi-scale feature fusion ability of the neck network and promote the lightweight of the model, part of the traditional convolution in the C2f module is replaced by spatial and channel reconstruction convolution (SCConv), and its structure is shown in Fig. 6. SCConv processes features orderly through spatial reconstruction unit (SRU) and channel reconstruction unit (CRU): SRU generates spatial attention weights, separates and reorganizes features to reduce spatial redundancy; CRU generates coordinate attention weights through multi branch convolution and global pooling to further aggregate key features and suppress redundant information, so as to enhance the ability of feature expression and reduce the number of parameters and computational complexity.

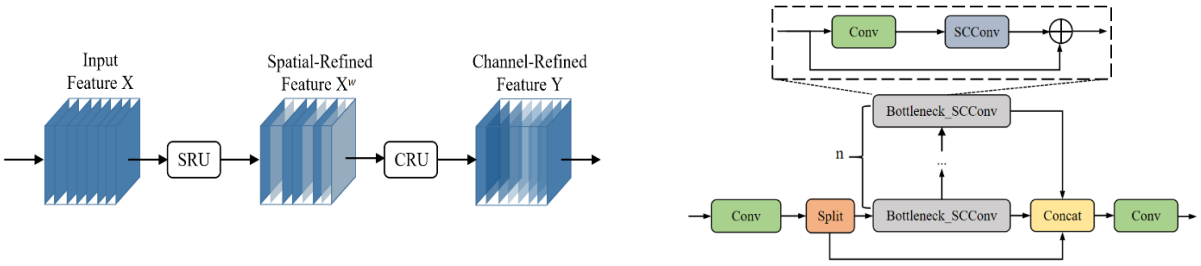


Fig. 6 – Structure of the C2f-SCConv Module

a) SCConv module; b) C2f-SCConv module

The C2f-SCConv module thus constructed inherits the C2f multi branch cross-scale interaction mechanism, introduces SCConv for feature decoupling and reconstruction, dynamically modeling feature correlation in spatial and channel dimensions, effectively compressing redundancy and enhancing key information. While improving the accuracy of small target detection, the design significantly reduces the model complexity and reasoning delay, achieves the balance between accuracy and efficiency, and is more suitable for real-time visual tasks with limited resources.

Detect-L module

In order to meet the urgent need of limited computing resources in agricultural visual real-time detection scene, a lightweight detection head, detect - L, was designed based on the YOLOv8 framework. The structure integrates depthwise convolution (DWConv), partial convolution (PConv) and simple attention module (SimAM) in turn, which can significantly reduce the number of parameters and computational complexity while ensuring the ability of multi-scale feature fusion. DWConv greatly reduces the computational burden by decomposing spatial filtering and channel fusion. PConv only performs spatial convolution on some channels to further reduce computational redundancy. SimAM can adaptively enhance the key feature regions and suppress the background interference according to the neural response energy without introducing any additional parameters. The structure is shown in Fig. 7.

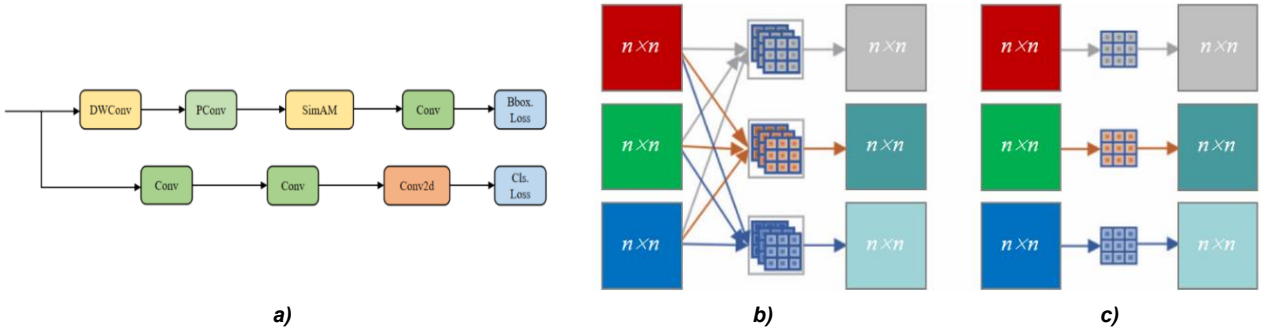


Fig. 7 – Detect-L Module Architecture and Convolution Operation Comparison
a) Detect-L Module; b) Standard convolution; c) DWConv

In particular, the SimAM attention mechanism used in this study directly derives the complete 3D attention weight of the feature map through the energy function, without adding trainable parameters, which is very efficient. Compared with the CA mechanism that focuses on establishing long-range channel dependence, SimAM can independently evaluate each spatial position in the feature map, so as to sensitively capture the local subtle features corresponding to small targets, effectively enhance key pixels and suppress noise.

The collaborative design of DWConv, PConv and SimAM not only maintains the detection accuracy, but also significantly improves the deployment applicability of the model in edge devices, providing a feasible solution for visual detection in resource constrained scenes.

RESULTS

Test environment configuration

The experiments in this study were conducted on the PyTorch deep learning framework using Python. The hardware configuration for testing comprised an Intel® Core™ i7-14700K CPU, 128 GB of RAM, and an NVIDIA GeForce RTX 3090 (24 GB) GPU, running on the Ubuntu 22.04 operating system. The software environment utilized PyTorch 2.3.1, Python 3.9.19, and CUDA 12.2. The detailed training parameters are listed in Table 1.

Table 1

Training Parameter Settings

Parameters	Values
Optimizer	SGD
Optimizer Momentum	0.937
Learning rate	0.01
Weight_decay	0.0005
Warmup_epochs	3
Warmup_momentum	0.8
Image size	640×640
Batch size	16
Epochs	200
Early stopping patience	50

Evaluation indicators

In order to comprehensively evaluate the performance of the model, this study selected key indicators of in-depth learning such as accuracy (P), recall (R), average precision (AP) and mean average precision (map). These indicators can quantitatively analyze the model detection effect from multiple dimensions, accurately reflect its performance in the target recognition task, and provide the basis for scientific and objective evaluation of model performance.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P \cdot (R) \mathrm{d}R \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{4}$$

where:

TP is the number of true positives;

FP is the number of false positives;

FN is the number of false negatives;

n is the number of classes.

In terms of model complexity evaluation, parameters, floating-point operations (FLOPs), and frames per second (FPS) are commonly used as core indicators. The number of parameters directly reflects the storage overhead and overfitting risk of the model, and its scale determines the model's expressive power and deployment difficulty. FLOPs measure the theoretical computational complexity of the model, directly affecting training and inference speed, and are a key consideration in lightweight design. FPS, as a performance metric during actual system operation, reflects the real-time processing capability of the model on specific

hardware, and is the direct basis for measuring whether an algorithm can meet the real-time requirements of actual scenarios. The combination of these three factors provides a comprehensive evaluation framework for the model in terms of efficiency, accuracy, and feasibility of deployment.

Detection capability comparison of attention mechanisms

To evaluate the effectiveness of the proposed Fast-CA module in bud orientation detection, it was compared with several representative attention mechanisms, including SE (Squeeze-and-Excitation) (*Ioannou et al., 2017*), CBAM (Convolutional Block Attention Module) (*Zhu et al., 2023*), and EMA (Efficient Multi-Scale Attention) (*Dong et al., 2024*). During the experiment, each attention module was integrated into the same location within the YOLOv8 network, while all other training parameters remained consistent to control variables. A comprehensive evaluation was conducted based on the number of parameters, computational cost (FLOPs), mean average precision (mAP), precision, and recall, mAP50 refers to the mean average precision evaluated with an IoU threshold of 0.5. with the results presented in Table 2.

Table 2

Detection Capability Comparison of Attention Mechanisms

Module	P/%	R/%	mAP50/%	Params/M	FLOPs/G
Fast-SE	87.8	89.4	93.7	**2.6**	7.0
Fast-CBAM	91.0	91.4	94.5	2.8	7.2
Fast-EMA	90.7	91.4	94.6	2.8	7.3
Fast-CA	**91.7**	**92.3**	**95.2**	2.7	**7.0**

As shown in Table 2, Fast-CA outperformed the compared methods in both precision and mAP, achieving improvements of 1.5, 0.7, and 0.6 percentage points over SE, CBAM, and EMA, respectively. Furthermore, Fast-CA also attained the highest scores in precision and recall, indicating its ability to enhance critical features while effectively suppressing background interference, thereby achieving superior overall detection performance.

Detection capability comparison of downsampling modules

To evaluate the effectiveness of the proposed LWDS module in bud orientation detection, it was compared with multiple downsampling modules, including YOLOV7-DS (the downsampling module in YOLOv7) (*Wang et al., 2023*), SPD-Conv (*Sunkara et al, 2022)*, and WaveletPool (*Williams et al., 2018)*. During the experiment, all training parameters except for the downsampling mechanism were kept strictly consistent to eliminate interference from unrelated variables. The evaluation was conducted based on the number of parameters, floating-point operations (FLOPs), mean average precision (mAP), precision, and recall, with the results presented in Table 3.

Table 3

Detection Capability Comparison of Downsampling Modules

Module	P/%	R/%	mAP50/%	Params/M	FLOPs/G
YOLOV7-DS	90.3	91.6	94.3	2.7	7.8
SPD-Conv	**91.7**	**92.3**	94.5	4.7	11.6
WaveletPool	90.4	90.8	94.1	**2.6**	**7.3**
LWDS	90.7	91.2	**95.0**	2.7	7.9

As shown in Table 3, LWDS demonstrates a significant advantage in mAP50 compared to YOLOV7-DS, SPD-Conv, and WaveletPool, achieving improvements of 0.7, 0.5, and 0.9 percentage points, respectively. The number of parameters in LWDS is comparable to that of YOLOV7-DS and is 2 M lower than SPD-Conv. In terms of FLOPs, LWDS is similar to YOLOV7-DS, significantly lower than SPD-Conv, and only slightly higher than WaveletPool. In summary, while maintaining low parameter and computational costs, LWDS achieves superior detection accuracy, demonstrating the effectiveness of this downsampling module.

Performance comparison of lightweight detection heads

To evaluate the effectiveness of the proposed Detect_L module in bud orientation detection, it was compared with several lightweight detection heads, including Detect_LADH, Detect_LSCD, and Detect_Seam.

Throughout the experiment, all training parameters except for the detection head were kept strictly consistent to eliminate interference from unrelated variables. The evaluation was conducted based on the number of parameters, floating-point operations (FLOPs), mean average precision (mAP), precision, and recall, with the results presented in Table 4.

Table 4

Performance Comparison of Lightweight Detection Heads

Module	P/%	R/%	mAP50/%	Params/M	FLOPs/G
Detect_LADH	89.7	90.0	94.5	2.4	5.7
Detect_LSCD	91.3	91.9	94.2	2.4	6.5
Detect_Seam	90.7	**92.1**	94.6	2.8	7.0
Detect_L	**91.4**	91.1	**94.8**	**2.4**	**5.5**

As shown in Table 4, Detect_L demonstrates a significant advantage in mAP compared to Detect_LADH (*Fan et al., 2024*), Detect_LSCD *(Ning et al, 2024)*, and Detect_Seam (*Sun et al., 2025*), with improvements of 0.3, 0.6, and 0.2 percentage points, respectively. In terms of parameter count, Detect_L is on par with Detect_LADH and Detect_LSCD, while being 0.4 M lower than Detect_Seam. Furthermore, its FLOPs are the lowest among all compared modules. Additionally, Detect_L also achieves excellent performance in both precision (P) and recall (R), indicating that it effectively enhances overall detection performance and stability while maintaining a lightweight design.

Ablation study

To evaluate the performance improvement of FSL-YOLO, comparative experiments were conducted stepwise based on the baseline YOLOv8 network, and the effectiveness of each improved module was analyzed. The comparative results are shown in Table 5.

Table 5

Ablation study results comparison

Base Model	Improved modules				P/%	R/%	mAP50/%	Params/M	FLOPs/G
	Faster-CA	C2f-SCConv	LWDS	Detect-L					
YOLOv8	×	×	×	×	88.4	90.9	93.7	3.0	8.1
	√	×	×	×	91.7	92.3	95.2	2.7	7.0
	×	√	×	×	91.5	92.5	95.3	2.8	7.7
	×	×	√	×	90.7	91.2	95.0	2.7	7.9
	×	×	×	√	91.4	91.1	94.8	2.4	5.5
	√	√	×	×	93.0	93.6	95.8	2.5	6.7
	√	×	√	×	92.3	92.6	95.6	2.3	6.9
	×	√	√	×	89.8	91.5	95.0	2.5	7.6
	√	√	×	√	92.3	**93.8**	95.9	2.1	4.1
	√	×	√	√	92.3	92.5	95.8	1.7	4.3
	×	√	√	√	91.2	92.0	95.1	1.9	4.9
	√	√	√	×	92.7	93.5	96.0	2.1	6.5
	√	√	√	√	**93.6**	93.7	**96.1**	**1.5**	**3.9**

Note: "×" indicates not using this module; √ "indicates the adoption of this module.

First, the original C2f module in the backbone network was replaced with the Faster-CA module integrated with the CA attention mechanism. This modification reduced the computational cost and parameter count to 86.4% and 90% of the baseline, respectively. Precision and recall increased by 3.3 and 1.4 percentage points, respectively, while mAP50 improved by 1.5 percentage points. These results indicate that the Faster-CA module not only reduces model complexity but also enhances feature extraction for small targets. Next, part of the standard convolution in the C2f module of the neck network was replaced with the SCConv module. This adjustment further reduced computational cost to 82.7% of the baseline and decreased parameter count by 16.7%.

Precision, recall, and mAP50 increased by 4.6, 2.7, and 2.1 percentage points, respectively. Although the multi-branch design of SCConv slightly increases computational overhead, it substantially improves feature representation capability. Subsequently, the LWDS was incorporated. The computational cost and parameter count were further compressed to 80% and 70% of the baseline, respectively. Precision, recall, and mAP50 improved by 4.3, 2.6, and 2.3 percentage points, with mAP50 reaching 96%. This demonstrates that LWDS effectively enhances multi-scale feature extraction while achieving model lightweighting. Finally, the Detect-L detection head was adopted for optimization. The computational cost and parameter count were significantly reduced to 48.1% and 50% of the baseline, respectively. Compared to the baseline, precision, recall, and mAP50 increased by 5.2, 2.8, and 2.4 percentage points.

In summary, by progressively integrating lightweight modules, the proposed FSL-YOLO model significantly reduces model complexity while steadily improving detection accuracy, achieving an effective balance between lightweight design and performance.

Detection capability comparison of different models

To evaluate the comprehensive performance of FSL-YOLO in the task of peanut sprout orientation detection, it was compared with mainstream lightweight detection models including YOLOv6n, YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and the RTDETR series. The results are presented in Table 6.

Table 6

Detection Capability Comparison of Different Models

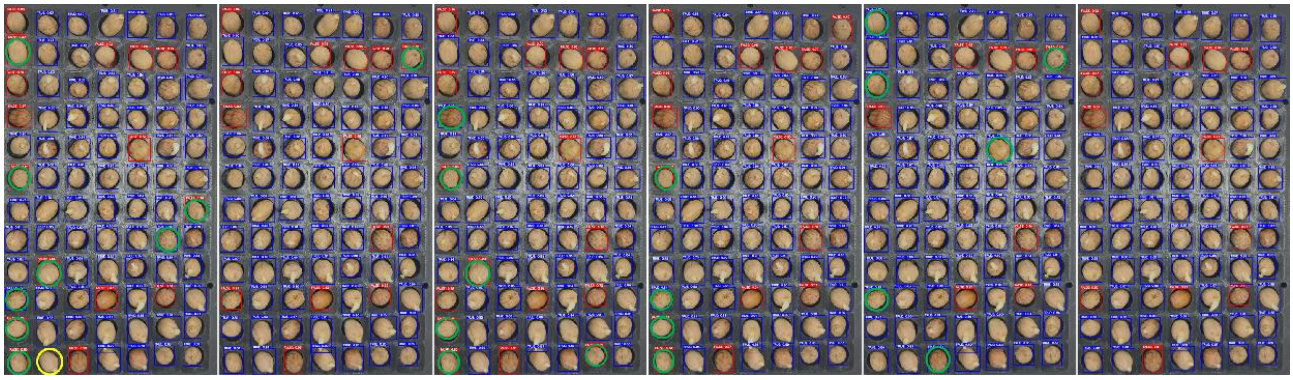
Module	P/%	R/%	mAP50/%	Params/M	FLOPs/G	FPS
YOLOv6n	89.7	89.5	94.5	4.2	11.5	62.27
YOLOv8n	91.6	92.7	94.7	3.0	8.1	**138.36**
YOLOv9t	92.2	93.6	93.6	1.7	6.4	85.22
YOLOv10n	90.1	92.2	94.7	2.7	8.2	118.39
YOLOv11n	91	90.1	94.0	2.6	6.3	123.30
RTDETR-l	86	86	88.1	32.0	103.4	30.32
RTDETR-resnet50	91.9	91.4	91.5	41.9	125.6	31.49
RTDETR - resnet101	91.2	91.4	91.6	60.9	186.2	27.41
FSL-YOLO	**93.6**	**93.7**	**96.1**	**1.5**	**3.9**	111.44

The comparative analysis shows that FSL-YOLO performs outstandingly in both lightweight design and detection accuracy: its parameter count is only 1.5 M, its computational cost is only 3.9 G, and its mAP50 reaches 96.1 %. Compared with other models, FSL-YOLO achieves the highest detection accuracy while maintaining the lowest parameter count and computational overhead. Specifically, compared to YOLOv6n, YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11n, its mAP50 is improved by 1.6, 1.4, 2.5, 1.4, and 2.1 percentage points, respectively. Compared with RTDETR-l, RTDETR-ResNet50, and RTDETR-ResNet101, the improvements are even more pronounced, reaching 8.0, 4.6, and 4.5 percentage points, respectively. Although the FPS of FSL-YOLO is slightly lower than that of lightweight models such as YOLOv8n, YOLOv10n, and YOLOv11n, it still remains at a relatively high level, which is sufficient to meet real-time detection requirements.

These results demonstrate that FSL-YOLO effectively enhances detection accuracy while significantly reducing model size, achieving a balance between lightweight design and performance optimization. Its overall performance is superior to that of the compared models.

Fig. 8 visually compares the detection performance of YOLOv6n, YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and FSL-YOLO in the task of peanut sprout orientation classification. It can be observed that the first five models exhibit varying degrees of missed and false detections, whereas the detection results of FSL-YOLO closely match the ground truth, with no obvious missed or false detections. Further analysis reveals that, in terms of error types, false detections are more frequent than missed detections. Regarding target characteristics, the models perform more accurately on longer, well-defined sprouts, but show poorer detection performance on shorter sprouts with less distinct features. This is presumably because lightweight models, in order to balance inference speed and parameter size, often compress network capacity and receptive field,

resulting in insufficient feature extraction capability for small, densely distributed targets, which in turn leads to localization deviations and misidentifications.



a) b) c) d) e) f)

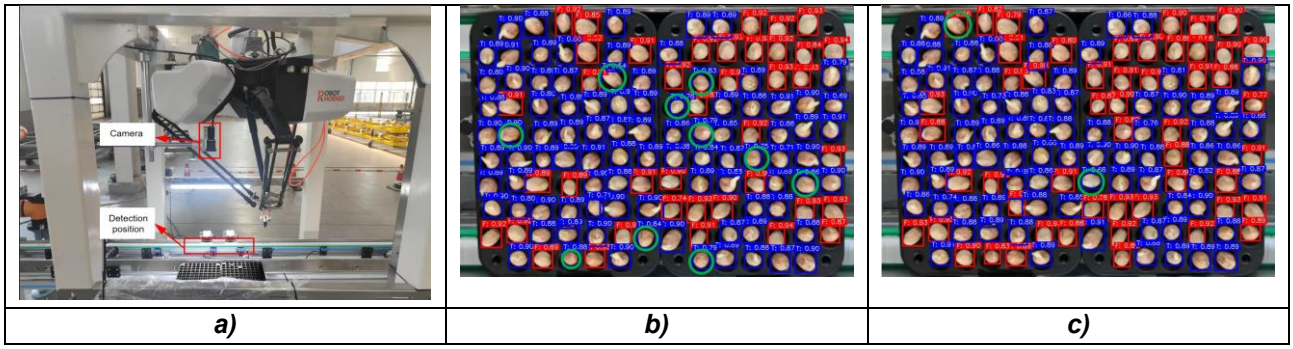
Fig. 8 – Comparison of Detection Results among Different Models

a) YOLOv6n; b) YOLOv8n; c) YOLOv9t; d) YOLOv10n; e) YOLOv11; f) FSL-YOLO

Note: The green circles indicate misdetected targets, while the yellow circles indicate missed targets.

Edge equipment deployment test

During the testing phase, to validate the actual detection performance of the improved algorithm on edge devices, this study deployed the YOLOv8n model and the proposed FSL-YOLO model on an industrial control computer running the Windows operating system for peanut bud orientation recognition tasks. The industrial computer is equipped with an Intel Core™ i7-12700 processor (2.1 GHz base frequency) and 16 GB of memory, with key hardware and software configurations consistent with the training environment. Model inference was performed using the ONNX Runtime framework (version 1.18.1). The test results and the bud orientation detection setup are shown in Fig. 9.



a) b) c)

Fig.9 – Bud Orientation Recognition Device and Test Results

a) Bud orientation recognition system; b) YOLOv8n; c) FSL-YOLO

As shown in Fig. 9, the device automatically detects the orientation of peanut buds through a bud-orientation identification system, and a flexible robotic gripper accurately removes those with bud-down orientation, retaining only the peanut buds that meet the directional requirement (bud-up). This process replaces the manual sorting step in traditional peanut sprout production, effectively promoting the automation.

The results indicate that the detection performance of YOLOv8n is notably inferior to that of FSL-YOLO, thereby verifying the superiority of the FSL-YOLO model. However, FSL-YOLO still exhibits false detections in some cases. Preliminary analysis suggests that this may be due to factors such as small bud targets, insufficiently prominent features, and interference from on-site reflections. This issue will be further analyzed and optimized in future research.

CONCLUSIONS

To address the multiple challenges of small target size, high density, and limited computational resources in the task of peanut sprout orientation classification within plug tray environments, this study proposes a lightweight improved model named FSL-YOLO.

Through systematic architectural optimization and module design, the model achieves a notable balance between detection accuracy and lightweight performance. Based on the comprehensive research conducted, the following conclusions can be drawn:

1. Model Architecture Improvements

This study introduces FSL-YOLO, a lightweight model based on YOLOv8, designed for detecting small-target peanut sprout orientation in dense scenes. The model enhances feature extraction by embedding the Faster-CA module in the backbone network, reduces computational complexity through the lightweight downsampling module (LWDS), improves feature robustness by incorporating SCConv in the neck network, and further optimizes parameter efficiency and detection accuracy using the Detect-L detection head. Compared to the baseline model, FSL-YOLO increases the mean average precision by 2.4 percentage points while reducing computational load and parameter count by 51.9% and 50%, respectively, effectively balancing detection accuracy with model efficiency.

2. Performance Comparison with Other Models

Under the same training environment and configuration, the improved FSL-YOLO demonstrates significant advantages compared to mainstream lightweight models such as YOLOv6n, YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11, RTDETR-l, RTDETR-ResNet50, and RTDETR-ResNet101. It achieves detection accuracy improvements of 1.6, 1.4, 2.5, 1.4, 2.1, 8.0, 4.6, and 4.5 percentage points, respectively, over these models, while maintaining the lowest parameter count and computational cost.

By effectively enhancing detection accuracy without compromising its lightweight design, FSL-YOLO delivers excellent overall performance, enabling precise peanut sprout orientation detection and providing a technical foundation for automated production in peanut sprout cultivation.

3. Future Work

Although the proposed model achieves satisfactory detection accuracy, further improvements in recall and precision are still possible. Future research plans to integrate hyperspectral imaging technology to enable simultaneous identification of mildew in peanut sprouts based on the current detection framework, thereby expanding the model's application value in automated production and advancing intelligent detection in peanut sprout cultivation.

ACKNOWLEDGEMENT

This study was supported by Shandong Province Agricultural Machinery R&D, Manufacturing, Promotion and Application Integration Pilot Project (NJYTHSD-202304), the Key R&D Program Project of Shandong Province (2023CXGC010702), and the Key R&D Program of Shandong Province (Major Science and Technology Innovation Project) (2025CXGC010807).

REFERENCES

[1] Bai, W.W., Zhao, X.N., Luo, N., Zhao, X.N., Luo, B., Zhao, W., Huang, S., Zhang, H. (2023). Study of YOLOv5-based germination detection method for wheat seeds (基于 YOLOv5 的小麦种子发芽检测方法研究). *Acta Agriculturae Zhejiangensis*, Vol.35, pp. 45-454. Beijing/China.

[2] Bao W.X., Xie W.J., Hu G.S., Yang X.J., Su B.B. (2023) Wheat ear counting method in UAV images based on TPH-YOLO (基于 TPH-YOLO 的无人机图像麦穗计数方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.39, pp. 155-161. Beijing/China.

[3] Chen, Z.Y., Gao, L.X., Charles C., Butts, C.L. (2017). Analysison technology status and development of peanut harvest mechanization of China and the United States (中美花生收获机械化技术现状与发展分析). *Transaction of the Chinese Society for Agricultural Machinery,* Vol.48, pp. 1-21. Beijing/China.

[4] Ding Z., Wang M., Hu B., Chen, Z., Dong, C. (2025). Impurity detection of premium green tea based on improved lightweight deep learning model. *Food Research International*, Vol.200. pp.115516. Amsterdam/Netherlands.

[5] Dong G.G., Chen X.K., Fan X.P., Zhou J.P., Jiang H. (2024) Detecting Xinmei fruit under complex environments using improved YOlOv5s (基于改进 YOLOv5s 的复杂环境下新梅检测方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.40, pp. 118-125. Beijing/China.

[6] Fan Z.P., Wu Y.Y., Lu W., Chen M., Qiu Z.G. (2024) Lg-YOLOv8: A lightweight safety helmet detection algorithm combined with feature enhancement. *Applied Sciences*, Vol.14, pp. 10141. Basel/Switzerland.

[7] Fang H.M., Xu Q.W., Chen X.G., Zhang Q.Y., Lu J.L. (2025) Cotton top bud recognition method based on DFR-SN-YOLO (基于 DFR-SN-YOLO 的棉花顶芽识别方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol 41, pp. 162-174. Beijing/China.

[8] Genze, N., Bharti., Grieb. J.S., Schuitheiss, S., & Grimm, D. (2022). Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant methods*, Vol.16, pp. 157. London/UK.

[9] Girshick R, Donahue J, Darrell T, Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 580-587. Columbus/ USA

[10] Gu, H.Y., Li, Z.H., Li, T., Li, T.H. Li, N., Wei, Z.C. (2024). Lightweight detection algorithm of seed potato eyes based on YOLOv5 (基于 YOLOv5 的马铃薯种薯芽眼轻量化检测算法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.40, pp. 126-136. Beijing/China.

[11] Hou, J.L., Fang, L.F., Wu, Y.Q., Li, Y.H., Xi, R. (2021). Rapid recognition and orientation determination of ginger shoots with deep learning (基于深度学习的生姜种芽快速识别及其朝向判定). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.37, pp. 213-222. Beijing/China.

[12] Ioannou Y., Robertson D., Cipolla R. (2017) Deep roots: Improving CNN efficiency with hierarchical filter groups. *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1231-1240. Honolulu/USA.

[13] Li J., Wen Y., He L. (2023) Scconv: Spatial and channel reconstruction convolution for feature redundancy, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6153-6162. Vancouver/ Canada.

[14] Li, M., Xiao, Y.Y., Zong, W.Y., Song, B. (2024). Detecting chestnuts using improved lightweight YOLOv8 (基于改进 YOLOv8 模型的轻量化板栗果实识别方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.40. pp. 201-209. Beijing/China.

[15] Ma, C.W., Zhang, H., Ma, X.M., Wang, J.L., Zhang, Y.S., Zhang, X.A. (2024). Method for the lightweight detection of wheat disease using improved YOLOv8 (基于改进 YOLOv8 的轻量化小麦病害检测方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.40. pp. 187-195. Beijing/China.

[16] Ning S.T., Tan F., Chen X., Li X.H., Shi H., Qiu J.K. (2024) Lightweight corn leaf detection and counting using improved YOLOv8, *Sensors*, Vol.24, pp. 5279. Basel/Switzerland.

[17] Redmon J, Divvala S, Girshick R, Farhadi, A. (2016). You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 779-788. Las Vegas/USK.

[18] Ren S., He K., Girshick R., Sun J. (2017) Faster R-CNN: Towards real time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol.39, pp. 1137-1149. New York/USA.

[19] Sunkara R., Luo T. (2022) No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 443-459. Porto/ Portugal.

[20] Sun Y., Guo H., Chen X. (2025) YOLOv8n-SSDW: A lightweight and accurate model for barnyard grass detection in Fields, *Agriculture*, Vol.15, pp. 1510. Basel/Switzerland.

[21] Tu, L.F., Peng, Q., Wu, X.N., Gu, J.W., Li, C.S. (2022). Automatic Monitoring System of Rape Seed Germination Rate under Dark Light Environment (暗光环境下油菜种子发芽率自动监测系统). *Seed*, Vol.41, pp. 133-137. Guizhou/China.

[22] Wang C.Y., Bochkovskiy A., Liao H.Y. (2023) YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 7464-7475. Vancouver/Canada.

[23] Wang, C., Wu, X.h., Zhang, Y.Q., Wang, W.J. (2022). Recognizing weeds in maize fields using shifted window Transformer network (基于移位窗口 Transformer 网络的玉米田间场景下杂草识别). *Transactions of the Chinese Society of Agricultural Engineering*. Vol.38, pp. 133-142. Beijing/China.

[24] Wang J.P., He M., Zhen Q.G., Zhou H.P. (2024) Camellia oleifera fruit harvesting in complex environment based on COF-YOLOv5s (基于 COF-YOLOv5s 的油茶果识别定位). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.40, pp. 179-188. Beijing/China.

[25] Wang, M., Li,Y., Meng, H., Chen, Z.W., Gui, Z.Y., Li, Y.P., Dong, C.W. (2024). Small target tea bud detection based on improved YOLOv5 in complex background, *Frontiers in Plant Science*. Vol.15. 1393138. Lausanne/Switzerland.

[26] Wang, J.Z., Gu, R.R., Sun, L., Zhang, Y. (2021). Non-destructive Monitoring of Plug Seedling Growth Process Based on Kinect Camera (基于 Kinect 相机的穴盘苗生长过程无损监测方法). *Transactions of the Chinese Society for Agricultural Machinery*, Vol.52, pp. 227-235. Beijing/China.

[27] Wang, N., Ning, C.C., Zhao, Z., Li, J.W., Wang, F., Ren, H.T., Zhan, K., Yu, Q.Y. (2022). Purification process by macroporous resin and antibacterial efficacy of resveratrol from peanut buds (花生芽白藜芦醇大孔树脂纯化工艺及抑菌功效). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.38, pp. 293-301. Beijing/China.

[28] Wang, N., Li, N., Yu, Q.Y., Ning, C.C. Wang, L.Y., Li, Z.N., Ren, H.T. (2021). Study on the change of nutrients during the germination process of peanut sprouts (花生芽菜发芽过程中营养物质变化规律研究). *Acta Peanuts*, Vol.50, pp. 61-67. Qingdao/China.

[29] Williams, T., Li, R. (2018) Wavelet pooling for convolutional neural Leaming networks, *International Conference on Representations*, pp. 1-12. La Jolla/ USA.

[30] Yin H.J., Wang B.L., Jing Y.G., Li J.Z., Wang P.L. (2024) Improved YOLOv7 method for counting watermelons in UAV aerial videos (基于改进 YOLOv7 的无人机航拍视频西瓜计数方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.40, pp. 124-134. Beijing/China.

[31] Zhang, W., Du, Y.F., Li, X.Y., Liu, L, Wang, L.Z, Wu, Z.K. (2024) Online Detection Method of Corn Kernel Quality Based on FSLYOLO v8n (基于 FSLYOLO v8n 的玉米籽粒收获质量在线检测方法研究). *Transactions of the Chinese Society for Agricultural Machinery*, Vol.55, pp 253-265. Beijing/China.

[32] Zhang, S.H., Shen, L., Song, L.J., Han, T.F., Song, Y.Y., Fang, Y.L. Su, B.F. (2023). Identifying and positioning grape compound buds using RGB-D images (基于 RGB-D 图像的葡萄复芽识别定位方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 39.pp. 172-180. Beijing/China.

[33] Zhang, W.Z., Zhang, H.Y., Liu, S.F., Zeng, Y., Mu, G.Z., Zhang, T.T. (2023). Detection of potato seed buds based on an improved YOLOv7 model (基于改进 YOLOv7 模型的马铃薯种薯芽眼检测). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.39. pp. 148-158. Beijing/China.

[34] Zhu, X.Y., Chen, F.J., Zheng, Y.L., Li, Z.Q., Zhang, X.W. (2023) Identification of olive cultivars using bilinear networks and attention mechanisms (融合双线性网络和注意力机制的油橄榄品种识别). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.39, pp. 183-192. Beijing/China.